A generalized panel data switching regression model

Malikov, Emir and Kumbhakar, Subal C.

Department of Economics, State University of New York at Binghamton, Department of Economics, St. Lawrence University

28 May 2014

A Generalized Panel Data Switching Regression Model

Emir Malikov*[1,2] Subal C. Kumbhakar[1]

[1]Department of Economics, State University of New York at Binghamton, Binghamton, NY
[2]Department of Economics, St. Lawrence University, Canton, NY

First Draft: March 20, 2014
This Draft: May 28, 2014

Abstract

This paper considers a generalized panel data model of polychotomous and/or sequential switching which can also accommodate the dependence between unobserved effects and covariates in the model. We showcase our model using an empirical illustration in which we estimate scope economies for the publicly owned electric utilities in the U.S. during the period from 2001 to 2003.

Keywords: Correlated Effects, Multinomial Logit, Nested Logit, Panel Data, Polychotomous, Selection

JEL Classification: C33, C34

*(Corresponding author) *E-mail address:* emalikov@binghamton.edu.

We would like to thank an anonymous referee for many helpful comments. We also thank Pablo Arocena for kindly providing us with the data used in the empirical illustration.

1 Introduction

Sample selection is a common problem in empirical work. Unobserved heterogeneity among units in the data poses a further challenge for practitioners. However, the increased availability of panel data and some recent developments in the literature have alleviated these challenges. In the case of strictly exogenous covariates, Wooldridge (1995), Kyriazidou (1997) and Rochina-Barrachina (1999) offer several ways to tackle both the selectivity and unobserved effects that are allowed to be correlated with covariates in the model. For a comparison of these methods, see Dustmann and Rochina-Barrachina (2007).

The above panel data models, as well as their generalizations that followed later [see Semykina and Wooldridge (2010) and the references therein], largely focus on *binary* sample selection. However, in many instances researchers face selection (or regime switching) of polychotomous and/or sequential nature. Examples include production technology studies of the industries which contain fully specialized, partly specialized and integrated firms, studies of higher education decisions and many others.

To fill this void, we contribute to the literature by considering a generalized panel data model of polychotomous switching which also allows for the dependence between unobserved effects and covariates in the model. The model we consider can be thought of as a generalization of a standard switching regression model. We show that Wooldridge's (1995) estimator can be readily extended to the case of polychotomous and/or sequential selection. For consistency, our method requires strict exogeneity of covariates conditional on unobserved effects. We showcase our model using an empirical illustration in which we estimate scope economies for the publicly owned electric utilities in the U.S. during the period from 2001 to 2003.

2 Model

Consider a generalized panel data switching regression model with correlated unobserved effects:

$$y_{it}^r = \begin{cases} \mathbf{x}_{it}^r \boldsymbol{\beta}^r + \alpha_i^r + u_{it}^r & \text{if} \quad D_{it} = r \\ - & \text{otherwise} \end{cases} \tag{2.1a}$$

$$D_{it}^{r*} = \mathbf{w}_{it}^r \boldsymbol{\gamma}_t^r + \xi_i^r + v_{it}^r \ , \qquad i = 1, \dots, N; \ t = 1, \dots, T; \ r = 1, \dots, R \tag{2.1b}$$

where $\mathbf{x}_{it}^r$ and $\mathbf{w}_{it}^r$ are $1 \times K_r$ and $1 \times L_r$ vectors of exogenous covariates (which may overlap)[1] with corresponding conformable parameter vectors $\boldsymbol{\beta}^r$ and $\boldsymbol{\gamma}_t^r$. (α_i^r, ξ_i^r) are individual-specific unobserved effects that are allowed to be correlated with right-hand-side covariates. The outcome variable y_{it}^r is observed only if the rth regime is selected. The regime selection (switching) is governed by a latent variable D_{it}^{r*} with observable categorical realizations: $D_{it} = r$ if the rth regime is selected. While the disturbances u_{it}^r and v_{it}^r are orthogonal to $(\mathbf{x}_{it}^r, \mathbf{w}_{it}^r)$, their distributions are however allowed to be correlated, namely $\mathbb{E}[u_{it}^r v_{it}^r | \mathbf{x}_{it}^r, \mathbf{w}_{it}^r] \neq 0$.

We first formalize the regime switching equation (2.1b). For convenience, define $\mathbf{x}_i^r \equiv (\mathbf{x}_{i1}^r, \dots, \mathbf{x}_{iT}^r)$ and $\mathbf{w}_i^r \equiv (\mathbf{w}_{i1}^r, \dots, \mathbf{w}_{iT}^r)$.

Assumption 1. For $i = 1, \dots, N$, $t = 1, \dots, T$ and $r = 1, \dots, R$, the conditional mean of unobserved effects ξ_i^r in a regime switching equation r is a linear projection on $\mathbf{w}_i^r$, i.e., $\xi_i^r = \mathbb{L}\left[\xi_i^r \,|\mathbf{w}_i^r\right] + a_i^s$,

[1]While our model does not require exclusion restrictions and can accommodate the case of $\mathbf{x}_{it}^r = \mathbf{w}_{it}^r$, in practice it is helpful to have some elements of $\mathbf{w}_{it}^r$ excluded from $\mathbf{x}_{it}^r$.

where $\mathbb{E}\left[a_i^s \,|\mathbf{x}_i^r, \mathbf{w}_i^r\right] = 0$. The composite error $e_{it}^r \equiv v_{it}^r + a_i^s$ is identically and independently distributed, conditional on $(\mathbf{x}_i^r, \mathbf{w}_i^r)$, with the type I extreme value distribution over i.

Specifically, we let the linear projection $\mathbb{L}\left[\xi_i^r \,|\mathbf{w}_i^r\right]$ take Chamberlain's (1980) form, i.e.,[2]

$$\mathbb{L}\left[\xi_i^r \,|\mathbf{w}_i^r\right] = \mathbf{w}_{i1}^r \boldsymbol{\delta}_{t1}^r + \cdots + \mathbf{w}_{iT}^r \boldsymbol{\delta}_{tT}^r \equiv \mathbf{w}_i^r \boldsymbol{\delta}_t^r \ . \tag{2.2}$$

Thus, our model allows for dependence between unobserved effects ξ_i^r and right-hand-side covariates $\mathbf{w}_{it}^r$. This formulation of correlated effects is essentially the one used in (Wooldridge, 1995, p.124).[3] One may alternatively permit $\mathbb{L}\left[\xi_i^r \,|\mathbf{w}_i^r\right]$ to take a more restrictive, but parsimonious, specification à la Mundlak (1978) which restricts $\boldsymbol{\delta}_{t1}^r = \cdots = \boldsymbol{\delta}_{tT}^r$ (e.g., Semykina and Wooldridge, 2010).[4] We also note that, unlike Wooldridge (1995) who assumes a normally distributed error in the selection equation, we assume the extreme value distribution, which is dictated by a polychotomous nature of the choice set.

The latent variable D_{it}^{r*} in (2.1b) can naturally be thought of as measuring an individual's propensity to select the regime r. Hence, the rth regime is said to be selected if and only if

$$D_{it} = r \quad \Leftrightarrow \quad D_{it}^{r*} > D_{it}^{j*} \quad \forall \quad j = 1, \ldots, R \ (j \neq r) \ . \tag{2.3}$$

While one can treat the regime switching as a system of $(R-1)$ dichotomous decision rules, we follow an alternative approach by considering the former in the random utility framework. That is,

$$D_{it} = r \quad \Leftrightarrow \quad D_{it}^{r*} > \max_{j=1,\ldots,R \ (j\neq r)} \left\{ D_{it}^{j*} \right\} \ . \tag{2.4}$$

After substituting for D_{it}^{r*} in (2.4) from (2.1b) and making use of Assumption 1, we let

$$\epsilon_{it}^r \equiv \max_{j=1,\ldots,R \ (j\neq r)} \left\{ \mathbf{w}_{it}^j \boldsymbol{\gamma}_t^j + \mathbf{w}_i^j \boldsymbol{\delta}_t^j + e_{it}^j \right\} - e_{it}^r \ . \tag{2.5}$$

From (2.5) it then follows that

$$D_{it} = r \quad \Leftrightarrow \quad \epsilon_{it}^r < \mathbf{w}_{it}^r \boldsymbol{\gamma}_t^r + \mathbf{w}_i^r \boldsymbol{\delta}_t^r \ . \tag{2.6}$$

Given that e_{it}^r is extreme value distributed, it follows that ϵ_{it}^r is multinomial logistically distributed over i with the corresponding marginal distribution $\Lambda_r(\cdot)$:

$$\Pr\left[D_{it} = r \,|\mathbf{x}_i^r, \mathbf{w}_i^r\right] = \Lambda_r\left(\mathbf{w}_{it}^r \boldsymbol{\gamma}_t^r + \mathbf{w}_i^r \boldsymbol{\delta}_t^r\right) = \frac{\exp\left(\mathbf{w}_{it}^r \boldsymbol{\gamma}_t^r + \mathbf{w}_i^r \boldsymbol{\delta}_t^r\right)}{\sum_j \exp\left(\mathbf{w}_{it}^j \boldsymbol{\gamma}_t^j + \mathbf{w}_i^j \boldsymbol{\delta}_t^j\right)} \ . \tag{2.7}$$

For some strictly positive monotonic transformation $\mathrm{J}_r(\cdot)$, condition (2.6) is equivalent to

$$D_{it} = r \quad \Leftrightarrow \quad \mathrm{J}_r(\epsilon_{it}^r) < \mathrm{J}_r\left(\mathbf{w}_{it}^r \boldsymbol{\gamma}_t^r + \mathbf{w}_i^r \boldsymbol{\delta}_t^r\right) \ . \tag{2.8}$$

We can now look at model (2.1) as a *binary* selection model, for each given regime r. That is, we can essentially replace the regime switching equation (2.1b) for each $r = 1, \ldots, R$ with its equivalent under Assumption 1:

$$\widetilde{D}_{it}^{r*} = \mathrm{J}_r(\mathbf{w}_{it}^r \boldsymbol{\gamma}_t^r + \mathbf{w}_i^r \boldsymbol{\delta}_t^r) - \mathrm{J}_r(\epsilon_{it}^r) \ , \tag{2.9}$$

[2] Clearly, $\boldsymbol{\delta}_{tt}^r$ is not identified here.

[3] The formulation of equation (2.1b) under Chamberlain's (1980) specification (2.2) is also equivalent to a reduced form of the following *dynamic* regime switching equation: $D_{it}^{r*} = \rho^r D_{it-1}^{r*} + \mathbf{w}_{it}^r \boldsymbol{\gamma}^r + v_{it}^r$.

[4] In this case, the linear projection in (2.2) is assumed to be a single index of the time averages of $\mathbf{w}_{it}^r$.

where $\widetilde{D}_{it}^{r*}$ is a transformed latent variable such that $D_{it} = r$ if and only if $\widetilde{D}_{it}^{r*} > 0$, i.e., condition (2.8) is satisfied. We follow Lee (1982, 1983) and consider $\mathrm{J}_r(\cdot) \equiv \Phi^{-1}[\Lambda_r(\cdot)]$, where $\Phi(\cdot)$ is the standard normal cdf. The advantage of such a transformation is that the random error $\mathrm{J}_r(\epsilon_{it}^r)$ in (2.9) is standard normal by construction, which would later enable us to make use of the truncated moments of the standard normal. Incidentally, the use of Lee's (1982, 1983) transformation as means of relaxing the normality in the selection equation has also been pointed out but not further developed in the panel data setting by Rochina-Barrachina (1999).

We next formalize the treatment of unobserved effects in the outcome equations of interest as well as the dependence between the two disturbances in (2.1a) and (2.9), where the latter enables us to correct for selectivity bias in the outcome equations.[5] For convenience, we define $\widetilde{\epsilon}_{it}^{\,r} \equiv \mathrm{J}_r(\epsilon_{it}^r)$.

Assumption 2. For $i = 1, \ldots, N$, $t = 1, \ldots, T$ and $r = 1, \ldots, R$:

(i) $\mathbb{E}\left[u_{it}^r \,|\mathbf{x}_i^r, \mathbf{w}_i^r, \widetilde{\epsilon}_{it}^{\,r}\right] = \mathbb{E}\left[u_{it}^r \,|\widetilde{\epsilon}_{it}^{\,r}\right] = \mathbb{L}\left[u_{it}^r \,|\widetilde{\epsilon}_{it}^{\,r}\right]$

(ii) $\mathbb{E}\left[\alpha_i^r \,|\mathbf{x}_i^r, \mathbf{w}_i^r, \widetilde{\epsilon}_{it}^{\,r}\right] = \mathbb{L}\left[\alpha_i^r \,|\mathbf{x}_i^r, \mathbf{w}_i^r, \widetilde{\epsilon}_{it}^{\,r}\right]$.

Assumption 2 states that the disturbance u_{it}^r is mean independent of $(\mathbf{x}_i^r, \mathbf{w}_i^r)$ conditional on $\widetilde{\epsilon}_{it}^{\,r}$. The latter holds if u_{it}^r and $\widetilde{\epsilon}_{it}^{\,r}$ are orthogonal to $(\mathbf{x}_i^r, \mathbf{w}_i^r)$, a standard assumption made in the sample selection models in the presence of strictly exogenous covariates. Unlike Wooldridge (1995), we also condition the expectation of u_{it}^r on $\mathbf{w}_i^r$, which is necessary because outcome and selection equations are permitted to have different covariates and non-zero cross-equation correlation between unobserved effects. Further, Assumption 2 does not impose any restrictions on temporal dependence of u_{it}^r or in the relationship between u_{it}^r and $\widetilde{\epsilon}_{it}^{\,r}$.
We specify the linear projection of u_{it}^r on $\widetilde{\epsilon}_{it}^{\,r}$ as

$$\mathbb{L}\left[u_{it}^r \,|\widetilde{\epsilon}_{it}^{\,r}\right] = \pi_t^r \widetilde{\epsilon}_{it}^{\,r}, \tag{2.10}$$

where π_t^r is time-varying to allow for temporal dynamics in the relationship between the two disturbances. A common alternative to (2.10) is the assumption of bivariate normality of the two errors which also implies linearity of the conditional mean of u_{it}^r (e.g., Lee, 1983). Our assumption is however less restrictive.

In order to account for correlated effects in outcome equations, Assumption 2 (ii) specifies the structure of unobserved heterogeneity. We follow Wooldridge (1995) and consider the following Chamberlain-type specification

$$\begin{aligned}\mathbb{L}\left[\alpha_i^r \,|\mathbf{x}_i^r, \mathbf{w}_i^r, \widetilde{\epsilon}_{it}^{\,r}\right] &= \mathbf{x}_{i1}^r \boldsymbol{\varphi}_1^r + \cdots + \mathbf{x}_{iT}^r \boldsymbol{\varphi}_T^r + \mathbf{w}_{i1}^r \boldsymbol{\omega}_1^r + \cdots + \mathbf{w}_{iT}^r \boldsymbol{\omega}_T^r + \psi_t^r \widetilde{\epsilon}_{it}^{\,r} \\ &\equiv \mathbf{x}_i^r \boldsymbol{\varphi}^r + \mathbf{w}_i^r \boldsymbol{\omega}^r + \psi_t^r \widetilde{\epsilon}_{it}^{\,r} \ .\end{aligned} \tag{2.11}$$

We now derive the selection bias corrected outcome equations. Taking the conditional mean of y_{it}^r from (2.1a) for each regime r, we obtain

$$\begin{aligned}\mathbb{E}\left[y_{it}^r \,|\mathbf{x}_i^r, \mathbf{w}_i^r, D_{it} = r\right] &= \mathbf{x}_{it}^r \boldsymbol{\beta}^r + \mathbf{x}_i^r \boldsymbol{\varphi}^r + \mathbf{w}_i^r \boldsymbol{\omega}^r + (\psi_t^r + \pi_t^r)\mathbb{E}\left[\widetilde{\epsilon}_{it}^{\,r} \,|\mathbf{x}_i^r, \mathbf{w}_i^r, D_{it} = r\right] \\ &= \mathbf{x}_{it}^r \boldsymbol{\beta}^r + \mathbf{x}_i^r \boldsymbol{\varphi}^r + \mathbf{w}_i^r \boldsymbol{\omega}^r + \rho_t^r \mathbb{E}\left[\widetilde{\epsilon}_{it}^{\,r} \,|\widetilde{\epsilon}_{it}^{\,r} < \mathrm{J}_r(\mathbf{w}_{it}^r \gamma_t^r + \mathbf{w}_i^r \boldsymbol{\delta}_t^r)\right] \ ,\end{aligned} \tag{2.12}$$

where we have used (2.10) and (2.11) in the first equality, and (2.8) and strict exogeneity of $(\mathbf{x}_i^r, \mathbf{w}_i^r)$ in the last equality. $\rho_t^r \equiv (\psi_t^r + \pi_t^r)$. Given that $\widetilde{\epsilon}_{it}^{\,r}$ is standard normal by construction, the expected value term in (2.12) equals the negative of the inverse Mills ratio, i.e.,

$$\mathbb{E}\left[\widetilde{\epsilon}_{it}^{\,r} \,|\widetilde{\epsilon}_{it}^{\,r} < \mathrm{J}_r(\cdot)\right] = -\frac{\phi\left[\mathrm{J}_r(\cdot)\right]}{\Phi\left[\mathrm{J}_r(\cdot)\right]} = -\frac{\phi\left[\mathrm{J}_r(\cdot)\right]}{\Lambda_r(\cdot)} \ , \tag{2.13}$$

[5] For a counterpart in Wooldridge (1995), see his Assumption 3′ on p.126.

where $\phi(\cdot)$ is the standard normal pdf.

The model can be consistently estimated in two stages. In the first stage, we estimate $\boldsymbol{\gamma}_t^r$ (and $\boldsymbol{\delta}_t^r$) via maximum likelihood as specified in (2.7) performed for each time period t separately. The obtained estimates $\widehat{\boldsymbol{\gamma}}_t^r$ are then used to compute the selection bias correction term. In the second stage, we consistently estimate the main parameters of interest $\boldsymbol{\beta}^r$ via pooled least squares performed on (2.12) that includes predicted inverse Mills ratios (for each regime r, separately).

Remark 1. One needs to account for the use of the predicted regressors in the second stage when computing standard errors for $\boldsymbol{\beta}^r$. Also, it may be of particular interest to conduct inference across equations for different regimes. We suggest following Newey (1984) and casting the model in a multiple-equation system GMM framework which permits the derivation of an asymptotic variance-covariance matrix for our two-stage estimator. Alternatively, paired (nonparametric) bootstrap can be employed.

Remark 2. For expository purposes, the covariates $\mathbf{w}_{it}^r$ and parameters $\boldsymbol{\gamma}_t^r$ in the regime switching equation (2.1b) are both assumed to be regime-varying. In practice, one needs to impose an identifying restriction on parameters to be regime-invariant for regime-varying covariates unless the latter vary with individuals only. That is, depending on the empirical application, the first stage is to be estimated via conditional, multinomial or mixed logit.

Remark 3. An obvious drawback of the regime switching formulation in (2.1b) is the independence of irrelevant alternatives (IIA) which may be hard to justify in a given application. The latter can be easily relaxed by reformulating the regime switching as a nested (sequential) process. For instance, consider a two-tier regime switching. Redefine equation (2.1b) as[6]

$$D_{it}^{jr*} = \mathbf{w}_{1,it}^j \boldsymbol{\gamma}_{1,t}^j + \mathbf{w}_{2,it}^{jr} \boldsymbol{\gamma}_{2,t}^{jr} + \xi_i^{jr} + v_{it}^{jr} \ , \qquad j = 1, \ldots, J; \ r = 1, \ldots, R_j$$

where $\mathbf{w}_{1,it}^j$ and $\mathbf{w}_{2,it}^{jr}$ are partitioned sets of covariates: those varying across the first-tier limbs indexed by j and those varying across the second-tier branches indexed by r, respectively.

Then, make a *modified* Assumption 1 where (an appropriately defined) v_{it}^{jr} can be assumed to be identically and independently distributed, conditional on $(\mathbf{x}_i^{jr}, \mathbf{w}_{1,i}^j, \mathbf{w}_{2,i}^{jr})$, with the *generalized* extreme value distribution over i. It is easy to show that the "nested logit" counterpart of (2.7) then takes the following form

$$\Pr\left[D_{it} = rj\right] = \frac{\exp\left(\mathbf{w}_{1,it}^j \boldsymbol{\gamma}_{1,t}^j + \mathbf{w}_{1,i}^j \boldsymbol{\delta}_{1,t}^j + \varrho^j I^j\right)}{\sum_{m=1}^J \exp\left(\mathbf{w}_{1,it}^m \boldsymbol{\gamma}_{1,t}^m + \mathbf{w}_{1,i}^m \boldsymbol{\delta}_{1,t}^m + \varrho^m I^m\right)} \times \frac{\exp\left(\left(\mathbf{w}_{2,it}^{jr} \boldsymbol{\gamma}_{2,t}^{jr} + \mathbf{w}_{2,i}^{jr} \boldsymbol{\delta}_{2,t}^{jr}\right) / \varrho^j\right)}{\sum_{k=1}^{R_j} \exp\left(\left(\mathbf{w}_{2,it}^{jk} \boldsymbol{\gamma}_{2,t}^{jk} + \mathbf{w}_{2,i}^{jk} \boldsymbol{\delta}_{2,t}^{jk}\right) / \varrho^j\right)} \ ,$$

where $I^j \equiv \log\left(\sum_{k=1}^{R_j} \exp\left(\left(\mathbf{w}_{2,it}^{jk} \boldsymbol{\gamma}_{2,t}^{jk} + \mathbf{w}_{2,i}^{jk} \boldsymbol{\delta}_{2,t}^{jk}\right) / \varrho^j\right)\right)$ is the so-called "inclusive value", and ϱ^j is the scale parameter which captures the dissimilarity between second-tier regimes and can be shown to equal $\left(1 - \mathrm{Corr}[v_{it}^{jr}, v_{it}^{jk}]\right)^{1/2}$. For more on the estimation of nested logit models, see Wooldridge (2010). The modified switching regression model can then be estimated under the remaining assumptions in two stages as above.

[6] We assume that all covariates and parameters are varying with branches for expository purposes only. However, when estimating the model in practice, appropriate identifying restrictions and normalizations are due.

3 Empirical Illustration

To showcase our generalized model, we estimate scope economies for an unbalanced sample of 117 electric utilities owned by local governments in the U.S. in 2001–2003. The data include firms of three types: (i) upstream – utilities that generate electricity, (ii) downstream – utilities that distribute electricity, and (iii) integrated – utilities that both generate and distribute electricity. All power generators use fossil fuel only.

In order to quantify scope economies, we first need to estimate production technologies for the industry. We employ the dual approach and estimate the underlying production technology using the cost function in which all covariates are exogenous as justified by economic theory. Given the nature of product and the government regulation, it is widely accepted that electric utilities treat output quantities, input (and output) prices as fixed.

We define the following two outputs: net electricity generated (y_1) and peak demand (y_2). The inputs are physical capital (x_1), fuel (x_2) and others (x_3) (including labor) with the corresponding vector of prices (w_1, w_2, w_3). Here we opt for a parsimonious specification to avoid multicollinearity and to conserve degrees of freedom given a small sample size. The price of capital (w_1) is the sum of the interest rate on long-term liabilities and the depreciation rate. We compute the fuel price (w_2) by dividing the fuel expenses by the fuel consumption measured in British thermal units (BTU). We follow Arocena et al. (2012) and proxy the price of other inputs (w_3), which includes labor and other operating expenses, by the state index of average wages for all employees from the U.S. Census Bureau. Lastly, the cost (C) is defined as the sum of capital, fuel and operating expenses, where the latter includes generation, distribution, administrative and general operation and management expenses, customer accounts, customer service and sales expenses.

Given the three types of utilities, not all firms produce both outputs and make use of all three inputs in their operations. The three technological processes can be summarized as follows.

$$\begin{aligned}
\text{upstream:} &\quad T^1 : \{(x_1, x_2, x_3) \to y_1\} \\
\text{downstream:} &\quad T^2 : \{(x_1, x_3) \to y_2\} \\
\text{integrated:} &\quad T^3 : \{(x_1, x_2, x_3) \to (y_1, y_2)\}
\end{aligned}$$

That is, there are many observations in the data in which upstream and downstream utilities report zero values for some combination of x_2, y_1 and y_2. The latter is a "zero-value observation" problem common to studies of electric utilities that estimate a common technology (cost function) for all types of utilities (e.g., Arocena et al., 2012).

However, the assumption of a common technology shared by utilities of all three types is quite unrealistic and unlikely to hold in practice. We relax it by allowing technologies to be type-specific and estimating the cost function for each of the three utility types separately: $C^1(w_1, w_2, w_3, y_1; \boldsymbol{\beta}^1)$ for upstream, $C^2(w_1, w_3, y_2; \boldsymbol{\beta}^2)$ for downstream and $C^3(w_1, w_2, w_3, y_1, y_2; \boldsymbol{\beta}^3)$ for integrated firms. Note that our approach does not suffer from the problem of having to deal with zero-value variables because they do not appear in the equations.

Lastly, we recognize that the utility type is likely to be a product of an endogenous choice made by firms, which we model explicitly. We condition the choice of the utility type on (the log of) regime-invariant total sales (S) and total revenues (R) to capture (exogenous to firms) demand.

We estimate the first stage via multinomial logit, where we restrict $\boldsymbol{\gamma}_t^1 = \mathbf{0}$ (upstream type) for the identification. We have explored relaxing the IIA assumption by formulating a (partly degenerate) nested two-tier selection: (i) specialized (upstream/downstream) vs. (ii) integrated utilities. We however have consistently failed to reject the null of multinomial logit, i.e., $\varrho^i = \varrho^{ii} = 1$.

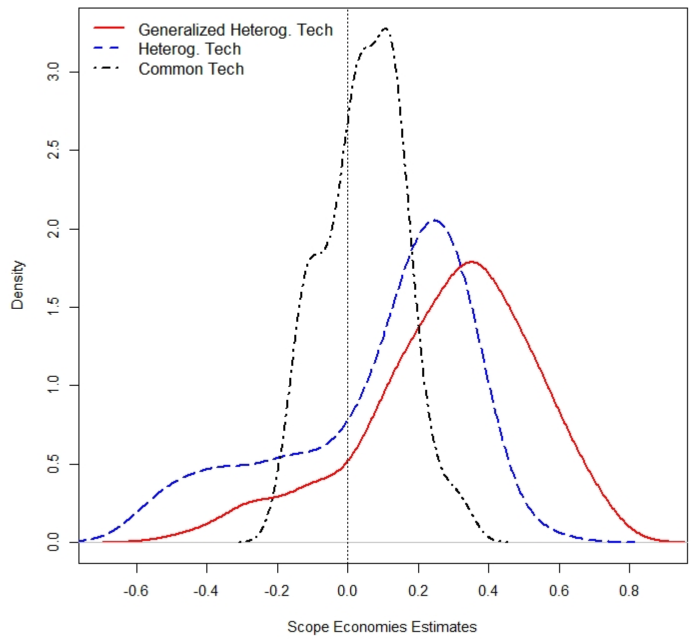

Figure 1: Kernel Densities of Scope Economies Estimates

In the second stage, we use the translog form for the dual cost function, onto which we impose the symmetry and linear homogeneity (by dividing input prices by w_3) restrictions. To conserve degrees of freedom, we define correlated effects in both stages using Mundlak's (1978) specification, as discussed in Section 2.

We use the fitted generalized model to compute scope economies exhibited in the electric utility industry. Unlike what is customarily done in the literature, we do not compute the statistic at some arbitrarily chosen data point (such as mean or median) but rather compute scope economies using actual data for integrated firms. The observation-specific scope economies are computed as

$$\mathrm{SE} = \frac{C^1(w_1, w_2, w_3, y_1; \boldsymbol{\beta}^1) + C^2(w_1, w_3, y_2; \boldsymbol{\beta}^2) - C^3(w_1, w_2, w_3, y_1, y_2; \boldsymbol{\beta}^3)}{C^1(w_1, w_2, w_3, y_1; \boldsymbol{\beta}^1) + C^2(w_1, w_3, y_2; \boldsymbol{\beta}^2)} \ . \tag{3.1}$$

Figure 1 plots kernel densities of the scope economies estimates from our generalized model (solid) as well as of the estimates obtained using two auxiliary (misspecified) models: (i) a model of heterogeneous technologies which estimated separate cost functions but ignores endogenous selection (dashed); and (ii) a model of common technology which fits one cost function for all types of utilities with zero-value observations set equal to 0.0001, as widely practiced in the literature (dot-dashed).

The figure suggests that the models that fail to account for technological heterogeneity and endogenous switching tend to underestimate scope economies: kernel densities based on both auxiliary models are to the left from that based on our generalized model. We attribute this to selectivity and misspecification biases present in the former models. We also formally test for the presence of endogenous switching via a joint Wald test of $\mathbb{H}_0: \ \rho_1^r = \cdots = \rho_T^r = 0$ for $r = 1, 2, 3$. The test rejects the null for upstream and downstream utilities with p-values less than 0.02. The corresponding median scope economies estimates from the three models (in the order used in the previous paragraph) are 0.31, 0.17 and 0.05. In particular, our generalized model predicts that integration

Table 1: Scope Economies Categories, %

	(I)	(II)	(III)
Scope Diseconomies	0	1.4	4.3
Scope Invariance	50.2	60.1	84.0
Scope Economies	49.8	38.5	11.7

NOTE: Model (I) – generalized model of heterogeneous technologies with endogenous selection; Model (II) – auxiliary model of heterogeneous technologies which ignores selection; Model (III) – auxiliary model of common technology.

of a power generator and a power distributor reduces cost by a median of 31%. The distribution of our estimates is consistent with findings reported in the literature (e.g., Kwoka, 2002).

However, kernel densities of the scope economies estimates in Figure 1 do not account for sampling errors associated with the estimation of models. Table 1 reports the breakdown of integrated electric utilities into three categories: Scope Diseconomies (SD), Scope Invariance (SI) and Scope Economies (SE). We classify a utility as exhibiting SD/SI/SE if its scope economies point estimate is statistically less/equal/greater than zero at the 95% significance level. Here we use a bootstrap two-stage, multiple-equation variance-covariance matrix obtained using 9,999 replications.

Based on our preferred generalized model, we find that as many as 50% of integrated electric utilities enjoy scope economies. The cost of the remaining half is invariant to the scope. The two auxiliary models however document a far worse picture, according to which 61% to 88% of integrated utilities exhibit scope invariance or, at worst, significant *dis*economies of scope.

4 Conclusion

We consider a generalized panel data model of polychotomous switching which also allows for the dependence between unobserved effects and covariates in the model. We contribute to the literature by extending Wooldridge's (1995) estimator to the case of polychotomous and/or sequential selection. The model is showcased using an empirical illustration in which we estimate scope economies for the publicly owned electric utilities in the U.S. during the period from 2001 to 2003.

References

Arocena, P., Saal, D. S., and Coelli, T. (2012). Vertical and horizontal scope economies in the regulated US electric power industry. *Journal of Industrial Economics*, 60(3):434–467.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225–238.

Dustmann, C. and Rochina-Barrachina, M. E. (2007). Selection correction in panel data models: An application to the estimation of females' wage equations. *Econometrics Journal*, 10(2):263–293.

Kwoka, J. E. (2002). Vertical economies in electric power: Evidence on integration and its alternatives. *International Journal of Industrial Organization*, 20:653–671.

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica*, 65(6):1335–1364.

Lee, L. (1982). Some approaches to the correction of selectivity bias. *Review of Economic Studies*, 49(3):355–372.

Lee, L. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(2):507–512.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, 14(2):201–206.

Rochina-Barrachina, M. E. (1999). A new estimator for panel data sample selection models. *Annales d'Economie et de Statistique*, 55-56:153–181.

Semykina, A. and Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157(2):375–380.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1):115–132.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, 2 edition.